

*Molding CNNs for text: non-linear,
non-consecutive convolutions*

Tao Lei, Regina Barzilay, and Tommi Jaakkola

Presented by Shih-Ming Wang
NLPLab, Institute of Information Science, Academia Sinica

07-5-2016

OUTLINE

INTRODUCTION

BACKGROUND

MODEL DESCRIPTION

EXPERIMENTS

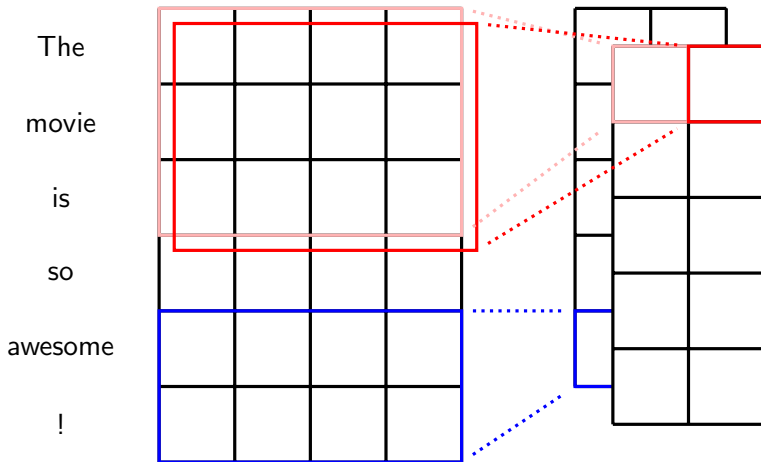
ERROR ANALYSIS

INTRODUCTION

MOTIVATION

- Deep learning & Convolution neural network (CNN) have led to success in many NLP problems
- Convolution operation is a **linear** mapping over **n-gram** vectors
- Target: **non-linear** operation over **non-consecutive** n-grams (e.g., “not that good”)

BACKGROUND



MODEL DESCRIPTION I

TENSOR-BASED FEATURE MAPPING

- Use outer product operation instead of linear combination
- Consider 2-gram (x_1, x_2) (row vectors) as example:

	Linear	Outer Product	3D case
Raw	$[x_1; x_2]$	$x_1^T \cdot x_2$	$x_1 \otimes x_2 \otimes x_3$
Dim(raw)	$2 \times d$	$d \times d$	$d \times d \times d$
Dim(Kernel)	$h \times 2 \times d$	$h \times d \times d$	$h \times d \times d \times d$
Output	$h \times 1$	$h \times 1$	$h \times 1$

,where $(x_1 \otimes x_2 \otimes x_3)_{ijk} = x_{1i} \cdot x_{2j} \cdot x_{3k}$

MODEL DESCRIPTION II

PARAMETER EXPLOSION

- Kernel T has $h \times d^n$ parameters for n-gram
- Solution: Decompose T in to sum of \bar{h} rank-1 tensors

	2D	3D
Dim(T)	$h \times d \times d$	$h \times d \times d \times d$
T'	$\sum_{i=1}^{\bar{h}} O_i \otimes P_i \otimes Q_i$	$\sum_{i=1}^{\bar{h}} O_i \otimes P_i \otimes Q_i \otimes R_i$

,where

$$O \in \mathbb{R}^{\bar{h} \times h}; P, Q, R \in \mathbb{R}^{h \times d};$$

$$O_i \in \mathbb{R}^h; P_i, Q_i, R_i \in \mathbb{R}^d$$

For simplicity, $\bar{h} = h$.

EXPERIMENTS

ERROR ANALYSIS